

# 합성곱신경망과 감성분석 기반의 가짜뉴스 탐지

이태원\*, 양영욱\*\*, 박지수\*\*\*, 손진곤\*<sup>1)</sup>

\*한국방송통신대학교 대학원 정보과학과

\*\*한신대학교 컴퓨터공학부

\*\*\*전주대학교 컴퓨터공학과

hep@knou.ac.kr

## Fake News Detection based on Convolutional Neural Network and Sentiment Analysis

Tae Won Lee\*, Yeongwook Yang\*\*, Ji Su Park\*\*\*, Jin Gon Shon\*

\*Dept. of Computer Science, Graduate School, Korea National Open University

\*\*Div. of Computer Engineering, Hanshin University

\*\*\*Dept. of Computer Science and Engineering, Jeonju University

### 요 약

가짜뉴스는 뉴스 기사 형식을 갖는 날조된 정보를 의미하며, 최근 모바일 인터넷 장치의 보급과 소셜 네트워크 서비스의 대중화로 온라인 확산이 가속화되고 있다. 기존 연구는 가짜뉴스의 탐지를 위해 뉴스의 주제, 부제목, 리드, 본문 등 뉴스 기사를 이루는 구성요소를 비롯하여 언론사, 기자, 날짜, 확산 경로 등의 메타 데이터를 대상으로 분석하였다. 그러나 뉴스의 제목과 본문 및 메타 데이터 등은 내용 수정이 쉬워, 다량의 데이터를 학습한 모델이라 하더라도 높은 정확도를 장기간 유지하기 어려울 수 있다. 이러한 문제점을 해결하기 위하여 본 논문은 합성곱 신경망을 이용해 문맥 정보를 분석하고 장단기 메모리 기반의 감성분석을 추가로 수행한다. 문맥 정보와 가짜뉴스 유포자가 쉽게 수정할 수 없는 감성 변화 패턴을 활용하여 성능이 개선된 가짜뉴스 탐지 모델을 제안한다.

### 1. 서론

가짜뉴스는 수많은 허위 정보 중에서도 뉴스 기사의 형식을 갖추고 있는 날조된 정보를 의미한다 [1]. 가짜뉴스가 갖는 정치, 경제, 사회적 영향력의 심각성에 세계가 주목하게 된 것은 비교적 최근의 일이다. 2016년 미국 대선에서 가짜뉴스가 많은 주목을 받았으며 트럼프 대통령의 당선에 중추적 역할을 했다는 의혹이 제기되기도 하였다[2]. 정치적 목적 외에도 가짜뉴스는 다양한 분야에서 영향력을 확장하고 있으며, 최근 이용이 급증하고 있는 모바일 인터넷은 소셜 붐에 의한 가짜뉴스의 광범위한 배포와 맞물려 가짜뉴스의 확산 속도가 가파르게 상승하게 된 원인이 되고 있다[3]. 가짜뉴스를 게시한 웹사이트의 광고를 중단하고[4], 언론의 고의적 허위 보도에 대해 손해배상 청구를 가능케 하는 언론중재법 개정안이 논의되는 등 가짜뉴스를 방지하기 위한 노력이 다양한 방면으로 시도되고 있지만, 가짜뉴스 확산의 효과적인 예방을 위해서는 가짜뉴스를 빠르고 정확하게 탐지하는 것이 무엇보다 중요하다.

최근 딥러닝[5]의 발전으로 자연어 처리 성능이 대폭 증가하였으며, 가짜뉴스 탐지의 성능 개선에 딥러닝이 효과가 있다는 것은 여러 선행 연구를 통해 입증되었다. 기존 연구는 뉴스 기사의 어휘 및 문맥 정보를 포함하고 있는 주제, 부제목, 리드(Lead), 본문 등을 주요 분석 대상으로 하여 가짜뉴스에서 공통으로 나타나는 특성을 추출하고자 하였다. 또한 부가적으로 언론사, 기자, 날짜, 확산 경로, 네트워크 트래픽 등의 메타 데이터(Metadata)를 참고하기도 하였다. 그러나 이 데이터들은 내용 일부를 쉽게 수정할 수 있어, 다량의 데이터를 분석한 모델이라 하더라도 높은 탐지 정확도를 장기간 유지하기 어려울 수 있다.

이러한 문제점을 완화하기 위하여 본 논문은 합성곱 신경망(Convolutional Neural Network; CNN)을 이용하여 가짜뉴스의 문맥적 특성을 분석하고, 장단기 메모리(Long Short-Term Memory; LSTM)를 통해 감성 변화 패턴을 추출한다. 감성 변화 패턴은 등장 순서에 의존적인 문장 성분별 감성 점수의 집합으로 가짜뉴스 유포자의 성격, 환경, 교육 및 가짜뉴스를 작성할 때 보이는 표현 습관을 반영하며

1) 교신저자 : 손진곤(jgshon@knou.ac.kr)

쉽게 바꿀 수 없는 특성이다. 감성 변화 패턴을 추출하기 위해 뉴스 기사를 토큰화(Tokenization)하고, 각 토큰에 감성분석을 적용하여 감성 점수를 할당한 후 LSTM을 사용한다. 마지막으로 문맥 정보와 감성 변화 패턴을 결합하여 뉴스 기사 사실 여부 식별에 활용한다.

## 2. 선행 연구

선행 연구에서 제안된 기법들은 뉴스 기사의 구성요소에 대한 정보를 새로운 관점에서 응용하는 방법과 외부 지식을 활용하는 방법으로 나눌 수 있다.

뉴스 기사의 구성요소에 대한 새로운 접근법인 토픽 모델링 활용 기법은 뉴스 기사로부터 추출한 10개의 이슈 그룹과 언론사 메타 데이터를 서포트 벡터 머신(Support Vector Machine; SVM)에 적용하고 학습 및 검증함으로써 공정성을 확보할 수 있게 하였다[6]. 또한 Fasttext와 Shallow-and-wide CNN을 이용하여 뉴스 기사 제목과 본문 간의 일관성을 분석함으로써 클릭 유도를 목적으로 하는 미끼 기사를 탐지하는 기법이 제안되기도 하였다[7]. 이러한 방법들은 가짜뉴스 유포자가 가짜뉴스의 본문과 메타 데이터를 의도적으로 수정할 경우, 분류 정확도가 쉽게 낮아질 수 있다는 한계점이 존재한다.

외부 지식을 활용하는 방법으로는 페이스북, 트위터와 같은 SNS와 뉴스 기사를 통합 활용하여 구축한 가짜뉴스 탐지 모델이 기존 모델과 비교했을 때 성능이 28% 향상됨을 보였고[8], 맥락적 단어 표현, 문장 내 엔티티 정보, 엔티티 정보에 대한 지식 그래프(Knowledge Graph)를 활용하고 어텐션(Attention)을 수행하여 가짜뉴스를 최종 분류하는 기법을 통해 데이터 세트 편향으로 인한 영향력을 완화하여 신뢰성을 높였다[9]. 외부 지식을 활용하는 접근 방법은 충분한 외부 지식 데이터 확보가 필수적이며, 신뢰할만한 지식 데이터 구축에 많은 시간과 비용을 투자해야 한다는 단점이 있다.

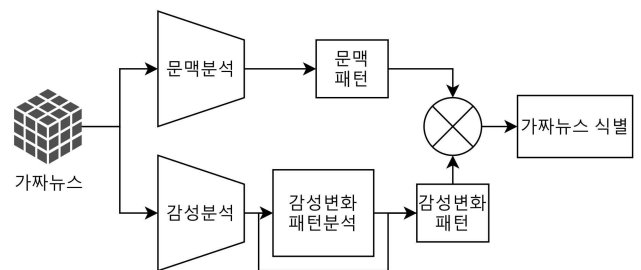
본 연구는 감성분석을 활용하여 데이터 구축에 큰 비용이 필요하지 않고 가짜뉴스 유포자가 쉽게 바꿀 수 없는 감성 변화 패턴을 추출한다. 감성분석은 텍스트 데이터에서 글쓴이의 주관적인 의견을 분석하여 긍정 또는 부정으로 분류하는 자연어 처리 기법으로 단어와 단어 간의 연결 관계에 따라 긍정과 부정이 레이블링 된 다량의 데이터와 순환 신경망(Recurrent Neural Network; RNN)을 활용하면 감성 사전의 구축 없이도 비교적 정확한 감성분석이

가능하다. 또한 본 논문에서는 순환 신경망의 단점인 기울기 소실(Gradient Vanishing), 기울기 폭주(Gradient Exploding) 문제를 완화한 LSTM을 적용하여 감성분석을 수행한다.

## 3. 가짜뉴스 탐지 모델

### 3.1 가짜뉴스 탐지 설계

선행 연구에 따르면 가짜뉴스는 허위 정보를 통해 의도적으로 독자의 불안을 부추기고, 독자에게 불확실성 지각을 통한 의심을 일으키는 등 긍정적 감성보다 부정적 감성과의 연관성이 높은 것으로 확인된다[10, 11]. 특히 재난 시 소셜 미디어에 의한 소문의 유포 및 확산에 관한 연구[12]에 의하면 부정적 감성은 가짜뉴스의 확산과 밀접하게 연관되어 있다. 본 연구에서는 감성분석을 통해 등장하는 순서에 의존적인 감성 변화 패턴을 추출하고, 부정적 감성에 높은 가중치를 부여함으로써 가짜뉴스의 특성에 최적화된 감성 특성을 추출한다. 또한 여전히 가짜뉴스 탐지에 있어서 중요한 통찰력을 제공하는 뉴스 기사의 어휘 및 문맥 정보를 분석하기 위하여 합성곱 신경망을 자연어 처리에 응용한 1D CNN[13]을 통합한다. 본 연구에서 제안하는 가짜뉴스 탐지 모델의 개요는 (그림 1)과 같다. <표 1>은 동작 과정을 설명하기 위한 가짜뉴스 데이터이며 A, B 모두 가짜뉴스로 레이블링이 완료되었다. A를 학습 데이터, B를 검증 데이터로 하는 테스트에서 B를 가짜뉴스로 분류하면 정확하게 탐지한 것으로 가정한다.



(그림 1) 가짜뉴스 탐지 모델

<표 1> 가짜뉴스의 예

A	미국, 영국 질병청에 보고된 백신 부작용 중증 사망자가 코로나 사망자 수를 훌쩍 넘기고 있다. 현 상황에서 백신은 답이 아닌 것으로...	가짜뉴스
B	경찰이 무력으로 백신 강제 접종할 수 있다. 문을 부수는 영상과 함께 한국에서도 경찰력을 동원한 백신 강제 접종이 이뤄질 것이라고...	탐지대상

3.2 가짜뉴스의 문맥 정보 분석

가짜뉴스의 문맥 정보 분석을 위해 CNN을 활용한다. 뉴스 기사를 워드투벡터(Word2vec) 기반의 벡터 집합  $x \in \mathbb{R}^{n \times d}$ 로 정의하고 합성곱 연산을 위한 커널을  $k^m \in \mathbb{R}^{l \times d}$ 로 정의하여 식 (1)에 따라 합성곱 연산을 수행한다. 식 (2)에 따라 커널  $k$ 에 해당하는  $t_n$ 으로부터 문맥 특성  $f^m \in \mathbb{R}^{n-l+1}$ 을 얻을 수 있다.

$$f = \sum_{i=1}^l \sum_{j=1}^d x(i, j)k(i, j) \quad (1)$$

$$t_n = \{x_n, x_{n+1}, \dots, x_{n+l-1}\}, \quad (2)$$

$$f_n^m = \text{Sigmoid}(t_n \otimes k + b)$$

<표 1>의 A, B 뉴스 기사에 CNN을 활용한 문맥 정보 분석을 적용하여 <표 2>의 정보를 추출할 수 있다. 모델이 A가 가짜뉴스임을 학습했어도 B를 가짜뉴스로 분류할 수 있는 근거가 부족하여 정확도 향상에 한계가 있다.

<표 2> 가짜뉴스의 문맥 정보

A	백신 부작용 사망자 > 코로나 사망자
B	백신 강제 접종 is 가능

3.3 감성 변화 패턴 분석

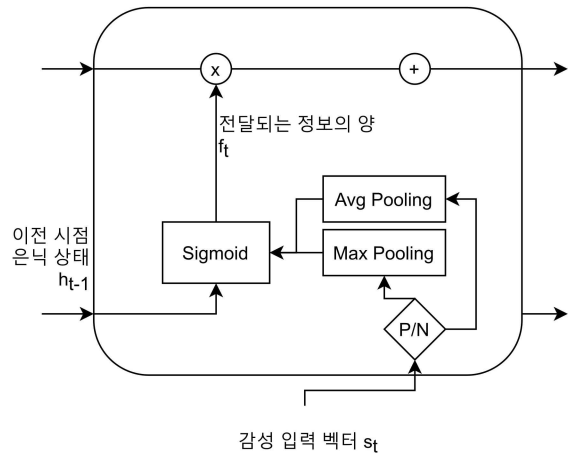
가짜뉴스  $x \in \mathbb{R}^{n \times d}$ 로부터 구두점을 기준으로 토큰화한 토큰 집합에 사전 학습(Pretrained)된 감성분석을 적용하여 문장별 감성 점수를 산출한다. 감성 점수의 집합  $s \in \mathbb{R}^{o \times e}$ 를 LSTM에 입력하여 등장 서에 의존적인 <표 3>의 감성 변화 패턴을 얻는다.

<표 3> 가짜뉴스의 감성 변화 패턴

A	부정 → 부정 → 부정 → 부정 → 부정 → 부정
B	부정 → 부정 → 부정 → 부정 → 부정 → 부정

LSTM에서 전달되는 정보의 양을 결정하는 삭제 게이트는 입력값  $s_t$ 와 이전 시점의 은닉상태  $h_{t-1}$ 에 시그모이드 활성화 함수를 적용하여 0과 1 사이의 값을 출력한다. 0에 가까울수록 정보가 많이 삭제되고, 1에 가까울수록 정보를 온전히 전달한다. 가짜뉴스가 부정적 감성과 밀접한 연관이 있는 것으로 나타나는 선행 연구에 따라 본 논문에서는 부정적 감성을 더 잘 전달할 수 있도록 삭제 게이트를 수정한

다. 삭제 게이트가 동작하는 시점에서 서브 샘플링(Sub-Sampling)을 통해  $t$ 시점의 감성 점수  $s_t$ 가 긍정적 감성일 경우 평균 풀링, 부정적 감성일 경우 최대 풀링을 적용하여 가짜뉴스 탐지 모델이 부정적 감성의 특성을 더 잘 추출할 수 있도록 한다. 가짜뉴스 탐지 최적화를 위해 수정된 삭제 게이트는 (그림 2)와 같다.

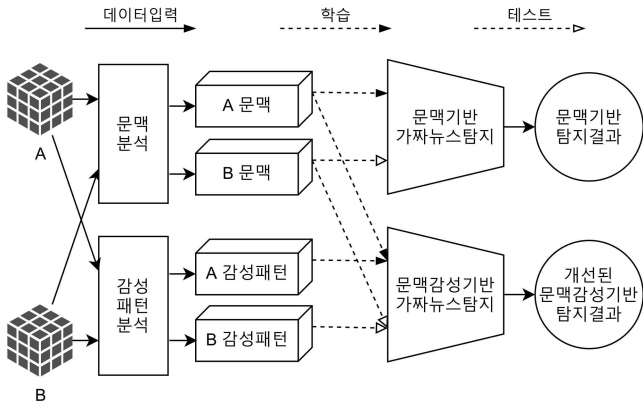


(그림 2) 가짜뉴스에 최적화된 LSTM 삭제 게이트

3.4 문맥 정보와 감성 변화 패턴의 결합

뉴스의 문맥 특성과 감성 변화 패턴의 결합값을 단일 벡터(Single Vector)로 결합하여 최종적인 가짜뉴스의 특성으로 정의하고 가짜뉴스 탐지에 활용한다. 문맥 정보와 감성 변화 패턴을 결합한 최종 결합값과 역전파(Back Propagation)를 이용한 딥러닝을 활용하여 가짜뉴스 탐지 모델을 학습한다.

<표 1>의 뉴스 기사 A, B는 코로나19 바이러스와 백신에 관한 가짜뉴스지만, 코로나19와 백신에 대해 다루고 있는 점을 제외하면 특성의 유사성을 찾기 어렵다. 따라서 문맥 정보만을 기반으로 A가 가짜뉴스임을 학습한 모델은 B를 가짜뉴스로 분류하기에 근거가 미흡할 수 있다. 그러나 감성분석을 적용함으로써 A, B 모두 코로나19 백신에 부정적인 뉴스 기사임을 알 수 있으며, 감성 변화 패턴이 일관성 있게 부정적인 것을 고려했을 때 부정적 감성을 의도적으로 형성하고 있음을 A, B의 공통적인 특성으로 추출할 수 있다. 가짜뉴스 탐지 결과를 개선하기 위해 문맥 정보와 감성 변화 패턴을 결합하는 과정은 (그림 3)과 같다.



(그림 3) 문맥 정보와 감성 변화 패턴의 결합

감성 변화 패턴은 뉴스 기사 간의 추가적 유사성을 얻을 수 있게 한다. 문맥 정보와 감성 변화 패턴을 기반으로 A가 가짜뉴스임을 학습한 모델이 B를 가짜뉴스로 분류할 수 있는 근거가 명확해지고, 이것은 탐지 정확도의 향상으로 이어질 수 있다.

**4. 결론 및 향후 연구**

본 논문에서는 가짜뉴스 탐지 성능을 개선하기 위해 문맥 정보 및 감성 변화 패턴을 추출하여 가짜뉴스 탐지에 활용하는 모델을 제안하였다. 감성 변화 패턴을 활용함으로써 가짜뉴스 탐지를 위한 새로운 통찰력을 얻을 수 있을 뿐만 아니라, 데이터에 대한 의도적인 조작을 어렵게 하여 딥러닝을 통해 학습된 가짜뉴스 탐지 모델의 정확도를 장기간 유지할 수 있다. 향후 연구에서는 사실 여부가 레이블링된 가짜뉴스 데이터 세트를 이용하여 기존 모델과의 비교 실험을 수행하고 성능을 검증하여 가짜뉴스에 나타나는 감성 변화 패턴과 가짜뉴스 탐지 성능 간의 연관성을 증명할 계획이다.

**참고문헌**

[1] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, Jonathan L. Zittrain, "The science of fake news", Science, 359, 6380, 1094-1096, 2018  
 [2] Allcott, Hunt, Matthew Gentzkow, "Social media and fake news in the 2016 election", Journal of economic perspectives, 31, 2, 211-36,

2017

[3] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu, "Fake news detection on social media: A data mining perspective", ACM SIGKDD explorations newsletter 19, 1, 22-36, 2017  
 [4] Wingfield, Nick, Mike Isaac, Katie Benner, "Google and Facebook take aim at fake news sites", The New York Times, 11, 12, 2016  
 [5] Krizhevsky, Alex, Ilya Sutskever, Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, 25, 1097-1105, 2012  
 [6] 윤태욱, 안현철, "텍스트 마이닝과 기계 학습을 이용한 국내 가짜뉴스 예측", Journal of Information Technology Applications & Management, 25, 1, 19-32, 2018  
 [7] Dong-Ho Lee, Yu-Ri Kim, Hyeong-Jun Kim, Seung-Myun Park, Yu-Jun Yang, "Fake news detection using deep learning", Journal of Information Processing Systems, 15, 5, 1119-1130, 2019  
 [8] 현윤진, 김남규, "뉴스와 소셜 데이터를 활용한 텍스트 기반 가짜 뉴스 탐지 방법론", 한국전자거래학회지, 23, 4, 19-39, 2018  
 [9] 조진욱, 김응범, 김한민, 손미애, "가짜뉴스 탐지를 위한 엔터티가 추가된 어텐션 네트워크 개발", 대한산업공학회 추계학술대회, 서울대학교, 2019, 2751-275  
 [10] Brashers, Dale E, "Communication and uncertainty management", Journal of communication, 51, 3, 477-497, 2001  
 [11] 최창식, "가짜뉴스에 대한 감정이 사실 확인 노력에 미치는 영향: 불확실성 관리 이론을 바탕으로", 커뮤니케이션 이론, 17, 1, 5-53, 2021  
 [12] Liu, Fang, Andrew Burton-Jones, Dongming Xu, "Rumors on Social Media in disasters: Extending Transmission to Retransmission", PACIS, 2014  
 [13] Kim, Yoon, "Convolutional Neural Networks for Sentence Classification", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746-1751, 2014